低环境可持续性降低环保人工智能的使用意愿(人工智能心理与治理专栏)[1]

魏心妮[1], 喻丰[2], 彭凯平[1]

[1]清华大学心理与认知科学系, 北京, 100084

[2] 武汉大学心理学系, 武汉, 430072

摘要 人工智能有助于生态环境治理和实现社会可持续发展，但它也正以惊人的速度消耗着能源，其产生的碳排放也影响着自然环境和人类生存。然而，暂未有研究关注人工智能产生的环境问题以及人类对此的反应，因此本研究探究了人机环境决策情境下，人工智能的可持续性对使用意愿的影响方式、原因和边界条件。预研究采用问卷调查并结合 ChatGPT 生成的态度词，考察了人们对人工智能环保系统的使用意愿和态度，发现人们的使用意愿较高且态度积极。研究 1 在 2 个子研究中分别操纵了人工智能可持续性的感知(有 vs.无)，并发现低可持续性组被试对人工智能的接受意愿更低，而且对国家开展人工智能研究的支持度也更低。研究 2 改变可持续性的操纵方式(低 vs.高)，再次通过实验重复了研究 2 的结果，并且发现道德而非能动是影响可持续性和接受意愿的中介机制。研究 3 探索了这一影响可能存在的边界条件，验证了个体亲环境态度的调节作用。研究结果为人工智能的社会治理提供了心理学依据，也为人工智能和可持续发展的关系提供了新的启示。

关键词 可持续性, 人工智能, 道德, 接受度

分类号 B849: C91

1 引言

人工智能(包括机器人、算法和模型)可以 “从经验中学习、适应新的输入，并执行类似人类的任务”(Duan et al., 2019)，因此有望帮助人类应对环境问题 (Nishant et al., 2020; Vinuesa et al., 2020) 。然而，人工智能与其他能耗设施一样，其研发、生产、使用和维护等各个生命周期都会消耗资源，并产生大量的碳排放 (Strubell et al., 2020) 。以自然语言处理训练模型为例，Strubell 等人(2020)估算，训练单个语言模型约产生 30 万公斤的二氧化碳，等同于从北京到纽约往返 125 个航班的碳足迹。这表明，如若要应用人工智能助力人类实现

[1] 收稿日期 2024-05-01
基金项目：国家社科基金青年项目(20CZX059)
通信作者：喻丰, E-mail: psychpedia@whu.edu.cn; 彭凯平, E-mail: pengkp@tsinghua.edu.cn

可持续发展目标(AI for sustainability)，其前提应当是人工智能是可持续的或者具备可持续的属性(sustainablity of AI, van Wynsberghe, 2021; Jay et al., 2024)。

但是，当前人工智能和可持续发展相关的研究存在积极发表偏向，忽略了人工智能的伦理及其可能对环境产生的潜在不利影响(Vinuesa et al., 2020)。更为重要的是，现有主题的研究主要集中于人工智能的技术层面(Verdecchia et al., 2023)，结论也多来自于计算机领域的会议论文(Schwartz et al., 2020)，而缺少从环境心理学的视角考察人类对于人工智能带来的负面环境影响有何行为反应(Al-Sharafi et al., 2023; Nishant et al., 2020)，且暂未有实证研究考察在环境治理场景中，人工智能可持续问题与人类使用意愿的关系。可持续意味着减少环境影响、提高资源效率，因此高可持续的人工智能有助于提升环境治理的工作表现，是符合社会环保规范和社会责任要求的，而这些与技术接受与使用统一模型(UTAUT)中个体的预期程度(performance expectancy)和社会影响(重要社会他人、规范的作用)对人工智能接受度影响的假设一致(Venkatesh et al., 2003)。所以，聚焦于环境治理情境下的人工智能系统，本研究旨在探究人工智能的可持续性如何影响人类的接受意愿，以期为人工智能的社会治理提供心理学研究证据。

1.1 可持续性与人工智能接受意愿

人工智能的可持续性，是指通过促进人工智能产品在整个生命周期(创意生成、训练、调整、实施和治理)的转变，在实现生态完整性和社会正义(van Wynsberghe, 2021)的同时保证碳足迹最低化(Pal, 2020)，即同时追求高准确性和高效能(Schwartz et al., 2020; van Wynsberghe, 2021)。然而，通过对计算机领域会议的 60 多篇文献进行分析，Schwartz 等人(2020)发现绝大多数人工智能的研究会优先考虑准确性(accuracy)而不是效能(efficiency)，因此这些研究实际上加速了“红色”人工智能(Red AI)的发展，即“以大算力买更好结果(buying stronger results by using massive compute)”，而这对环境是不友好的、有威胁的。

尽管暂未有实证研究揭示人工智能的可持续性和接受意愿之间的关系及影响机制，但以往研究为二者关系提供了部分间接证据。具有可持续属性的产品展示出生产者将资源用于解决环境问题和提高社会福祉的良好品质(Chen et al., 2020)，而选择绿色或环保产品则体现出个体对环境的关注和亲社会的动机(Schwartz & Loewenstein, 2020)，因此高可持续性能够提高消费者对产品的满意度和购买意愿(Hernandez et al., 2015; Grazzini et al., 2021) 。对于人工智能而言，可持续性可能影响人类对人工智能产品的态度 (Gansser & Reich, 2021)。而且，相较于低可持续性的人工智能，高可持续性的人工智能对环境更友好、对人类威胁更小(Schwartz et al., 2020) ，因此人们对高可持续的人工智能接受更高。根据技术接受与使用统一模型，当意识到高可持续性的人工智能对环境保护有积极作用时(即高结果预期)，人类对其接受度会更高，相反则对低可持续性的人工智能接受度更低。因此，本研究提出假设：可持续水平越低，人们对人工智能的接受意愿就越低。

1.2 道德能动性的中介作用

人工智能的发展应以增进人类福祉为目标，且要符合人类的价值和伦理道德(Djeffal et al., 2022)。具体在环境保护方面，人工智能不仅应当促进绿色发展[2]，还应当像人类一样遵守环境友好、资源节约的要求(Gansser & Reich, 2021; Siala & Wang, 2022)。因此，当人工智能违反道德规范时，人们会应用道德基础准则对其产生道德感知和判断(Banks, 2021; Maninger & Shank, 2022; Wilson et al., 2022)。这表明，人们把人工智能视为具有自主性的道德主体(moral agent)，而且认为其应当承担一定的道德责任(喻丰, 许丽颖, 2018; Bigman et al., 2023; Formosa & Ryan, 2021; van Wynsberghe & Robbins, 2019)，即人们会对人工智能的道德能动性产生感知和评价。道德能动性(moral agency)是指个体具有在道德决策中发挥主动作用的属性或能力，包括自主选择道德行为、评估道德情境、承担道德责任等(Banks, 2019; Formosa & Ryan, 2021)。道德能动性包括道德与能动两个方面 (喻丰, 2020; Floridi & Sanders, 2004)，且这两个维度又构成了人们感知某个人类或者机器是否具备道德能动性的必要和充分条件(Floridi & Sanders, 2004)。

道德是指主体运用道德账户(Lakoff, 1995)、道德准则(Haidt & Joseph, 2004)等辨别对和错的能力或系统 (Haidt, 2001)。以往研究表明，环保问题实质上是道德问题(Feinberg & Willer, 2013; Jia et al., 2017)，因此节约资源、绿色消费等行为是道德行为(Braun Kohlová & Urban, 2020)，而浪费资源和破坏环境等是不道德的行为(Bretter et al., 2023; Krettenauer, 2017)。低可持续性的人工智能不仅消耗资源还会产生大量碳排(Dhar, 2020)，这违反了道德基础理论中关爱和纯洁的道德准则 (Haidt & Joseph, 2004)。因此，人们认为复杂人工智能系统应该因其造成的环境破坏而受到道德指责(Kneer & Stuart, 2021)。这说明，相比于高可持续性人工智能，人们感知到低可持续的人工智能的道德水平可能更低。

能动是指人工智能可以自主地执行自我导向的行为(Himma, 2009)，这意味着其行动支配力从人类制造者转移到机器自身(Gunkel, 2012)。人类对行为者能动性的感知主要取决于其行为是否具有自主性(autonomy)和意图性(intentionality, 喻丰, 许丽颖, 2018; Sullins, 2006)。其中，自主表现为不依赖于其他主体而由自我内部控制的行动或运转(Deci & Ryan, 1987; Gunkel, 2012); 意图则主要表现为做出有意行为的自由意志，即行为主体是蓄意为之的而非依赖于远程程序或条件设定(Banks, 2019) 。尽管人们希望人工智能承担一定的道德责任，但是人们感知到的人工智能的能动水平，例如心智能力(沟通和思考)、选择能力和行为意图性影响了人们在多大程度上将其视为道德主体，以及在违反道德准则后人工智能应当受到多大程度的责备 (Maninger & Shank, 2022; Monroe et al., 2014)。低可持续性的人工智能是高耗能和高碳排放的，这意味着人工智能本身对能源的利用能力较低(Schwartz et al., 2020)，而且人们可能将这种结果归因于人类程序的设定，而非是人工智能自身有意为之(Moor, 2006)。所以，能动有时候也常用其反面作为指标，即依从(dependency), 其主要通过

[2] 2023-9-18 取自 https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html

评估人工智能缺少能动性和依赖程序或人类控制的程度来衡量(Bank, 2019)。由此，我们推测，人们感知到的低可持续性的人工智能依从性更高，进而降低人们对低可持续人工智能的接受意愿。

综上所述，本研究提出假设 2：

H2a：道德是影响可持续性和人工智能接受意愿的中介变量；

H2b：依从是影响可持续性和人工智能接受意愿的中介变量。

1.3 亲环境态度的调节作用

个体对人工智能可持续性问题的感知受到自身亲环境态度的影响(Gansser & Reich, 2021)。亲环境态度(pro-environmental attitude)指个体对环境或对环境问题的关心(Gifford & Sussman, 2012)，其主要思想是：地球的承载能力是有限的，生态系统需要保持平衡，而且人是自然界的一部分，所以经济发展不能以牺牲生态环境为代价(Dunlap et al., 2000)。以往研究表明，有志于保护、改善环境以及重视可持续性的人对环境友好技术的态度更开放，并认为这些技术是非常有用的(Averdung & Wagenfuehrer, 2011; Park et al., 2017)。这是因为，人们认为节能对环境是有益的，是环境责任感的体现，而且即使在不考虑个人经济利益的情况下，环境友好技术也被感知为是有用的(Mert et al., 2008)。这表明，个体的亲环境态度越强烈，而人工智能技术对环境越友好时，个体对人工智能的接受意愿就越高。据此，本研究提出研究假设 H3a：亲环境态度越强烈，人们对低可持续性的人工智能的接受意愿就越低。

同时，环保与个体的道德信念有关(Feinberg & Willer, 2013; Jia et al., 2017)，而且可能是导致人们环境问题态度两极分化的重要原因(McCright & Dunlap, 2011)。保护环境是社会规范和道德准则的基本要求(Farrow et al., 2017)，所以破坏环境的行为被感知为是不道德的(Graham et al., 2011; Tetlock, 2002)。为了减缓厌恶情绪，人们可能做出更多有益于环境的行为来达到道德洁净和平衡的目的(Johnson & Ahn, 2020)。因此，亲环境态度越强烈，个体对低可持续性的人工智能会产生更多的道德厌恶情绪和更低的道德评价，从而使得人们可能通过拒绝使用人工智能来消除消极的心理体验。并且，基于技术接受与使用统一模型的观点(Venkatesh et al., 2003)，低可持续的人工智能资源消耗更多且碳排放更多，而这会降低个体对人工智能有用性的预期程度，并感知到这与社会环保规范和社会责任是相违背的，由此导致人们对低可持续的人工智能的接受意愿也更低。所以，本研究提出假设 H3b: 亲环境态度越强烈，人们感知到的低可持续性人工智能的道德水平越低，从而降低人们对人工智能的接受意愿。

1.4 研究概览

为验证前述假设，我们开展了 4 个递进的实证研究。预研究通过 ChatGPT3.5 随机生成环保人工智能系统的态度测量词并开展问卷调查，以了解人们对人工智能环保系统的态度和接受意愿。研究 1 通过 2 个子研究操纵被试对人工智能可持续性的感知(有 vs.无)，以探

究可持续性对接受意愿的影响；研究 2 操纵被试对可持续性(高 vs.低)的感知，再次检验人工智能的可持续性和接受意愿之间的关系，并探究二者关系是否受到道德能动性的中介作用。接着，研究 3 通过实验探索了影响二者关系的边界条件，即亲环境态度对可持续性和接受意愿关系的调节作用。

2 预研究

由于人工智能和可持续发展相关的研究主要集中于模型训练阶段(Verdecchia et al., 2023)，暂缺少研究从环境心理学角度实考察人类对人工智能参与环境决策的心理和行为反应(Nishant et al., 2020)。因此，在正式研究前，通过 ChatGPT 生成若干态度形容词，以调查人们对于人工智能参与环境决策的基本态度，并为正式研究提供经过验证的、直接的测量工具。

2.1 被试

通过 Credamo 随机招募到 251 名被试，2 名被试因未通过注意力筛查而被剔除，剩下有效被试 249 名(其中男性 99 名，女性 150 名; 18~62 岁, M_{age} = 30.09 岁，*SD* = 7.69)。

2.2 研究过程

知情同意后，被试填写个人基本信息，包括年龄、性别以及受教育水平。接着，参考 Haesevoets 等人(2021)的研究，向被试呈现两段人工智能环保系统相关的信息，要求他们在阅读过程中想象自己是一名社区管理员。具体阅读信息如下：

请想象一下，目前你是一名社区管理者。作为管理者，你的日常工作决策通常会对社区居民的生活产生直接的影响。特别是社区的环境保护和可持续发展相关的问题，你必须就社区范围内每个居民小区的垃圾处理问题及时地做出检查、监督和教育等决定。再比如，社区辖区内垃圾处理站的设置、社区志愿协管员的工作安排等也是你需要考虑和决策的问题。

近年来，越来越多地组织开始利用自主人工智能(AI)来帮助解决环境问题。其中，人工智能环保系统 AECOS 是一种专门解决环保问题的智能系统，它使用算法来推断特定问题的规则，并根据社区各个垃圾处理站的运营情况、社区协管员的工作表现和时间安排进行自动评分，以保证做出的最佳决策。

阅读完毕后，被试回答“在多大程度上，你会向上级部门建议采购并使用人工智能环保系统 AECOS 来帮助社区处理环境问题?”(1=非常不可能, 7=非常可能)，以测量被试对人工智能系统的接受程度。同时，测量被试对人工智能环保系统的态度评价。由于暂未有研究关注人类对环保领域人工智能的反应，因此在通过人工智能系统 ChatGPT 3.5 生成若干态度评价词。在对话框中使用英文输入“请列举出 100 个用以描述人们对于将人工智能应用于解决环境的态度形容词，然后根据词语的意思将其进行分类。” 随后，ChatGPT 按照指令呈现了目标词，并将词语分成了“乐观的”(optimistic, 16 个)、“关心的”(concerned, 17 个)、“忧虑的”(apprehensive, 16 个)、“矛盾的”(ambivalent, 7 个)、“支持的”(supportive, 15 个)、“消极

的”(negative, 16 个)和“满怀希望的”(hopeful, 8 个)七个不同类别。然后，继续由 ChatGPT 从前述的七个类别中随机选择 14 个形容词，并将其生成为人工智能环保系统态度测量条目，其中积极态度词($\alpha = 0.80$)：激动的、有信心的、好奇的、有耐心的、接受的、赞同的、有追求的、先进的；消极态度词($\alpha = 0.82$)：怀疑的、不信任的、冷嘲热讽的、矛盾的、抵制的、有敌意的。在测量中，采用李克特七点计分，被试报告自己对各个条目的赞同程度(1 = 非常不同意, 7 = 非常同意)。

2.3 结果与分析

配对样本 T 检验结果表明，被试在积极态度词(M = 5.84, SD = 0.63)的评分显著高于消极态度词(M = 2.67, SD = 0.63)，$t(248) = 44.04$, $p < 0.001$, $d = 1.14$, 95% CI [3.03, 3.17]。总体而言，参与研究的被试对于人工智能环保系统的接受程度较高(M = 5.88, SD = 0.81)，并且接受度与积极态度显著正相关($r = 0.57$, $p < 0.001$)，与消极态度显著负相关($r = -0.41$, $p < 0.001$)。进一步地分析被试对人工智能接受度评分的情况，结果发现在非常不可能(1)到非常可能(7)的七个答案选项中，无被试选择非常不可能(1)和比较不可能(2)，6 名被试选择有点不可能(3)或持中立态度(4)。但有 243 名被试(97.59%)表示自己可能(5~7)接受人工智能参与社区环保治理工作，其中 21.7% (54 名)被试表示非常可能(7)向上级部门建议采购并使用人工智能环保系统 AECOS。这表明，人们对于使用人工智能解决环境问题持积极态度。

综上所述，预研究结合 ChatGPT 生成的态度词条目对被试进行随机调查，结果发现人们对于将人工智能应用于环境决策总体持以积极态度，且有较高的使用意愿。

3 研究 1

尽管预研究发现人们对人工智能环保系统的态度积极且使用意愿较高，但这可能是因为研究中使用的实验材料仅展现了人工智能对环境治理工作来带的便利性，而未考虑到人工智能对环境的带来的消极影响。因此，研究 1 在实验中向被试同时呈现了人工智能在环境治理场景中产生的积极和消极结果，并通过测量被试对国家开展人工智能研究的态度，来探究人们对人工智能的态度和使用意愿。根据研究假设 H1，人们对低可持续性人工智能的使用意愿更低。研究 1 通过 2 个子研究，并采用实验法来检验这一假设。

3.1 研究 1a

3.1.1 被试

根据 G*Power 3.1 的计算结果可知，被试量最少需为 172 才能使单因素两水平被试间实验达到中等效应量($f = 0.25$, power = 0.90)。通过 Credamo 随机招募到 241 名被试，其中 1 名被试未通过注意力筛查及 1 名被试的年龄小于 18 岁而被剔除，剩下有效被试 239 名(男性 93 名, 女性 146 名; 18~66 岁, M age = 30.98 岁, SD = 8.20)。

3.1.2 研究程序

在经过知情同意程序后，被试首先报告自己的基本人口学信息，以及所从事的职业与人工智能的相关程度(1=非常低, 7=非常高)。为操纵被试对人工智能可持续程度的感知，参

考 Strubell 等人(2020)对运用自然语言处理模型处理任务过程中所产生碳排放量的估算，并使用网络碳足迹计算器[3]计算了达到碳中和需要付出的努力。首先，所有被试随机进入到低可持续性人工智能组(n = 119)或控制组(n = 120)，并阅读到以下内容：“近年来，越来越多地组织开始利用自主人工智能(AI)来帮助解决环境问题。其中，人工智能环保系统 AECOS 就是一种专门解决环保问题的智能系统，它使用算法来推断特定问题的规则，并根据社区各个垃圾处理站的运营情况、社区协管员的工作表现和时间安排进行自动评分，以保证做出的最佳决策”。然后，控制组被试直接进入到操纵检查和因变量测量任务中。低可持续性组被试则继续阅读如下内容 “但是，人工智能本身也是重要的碳排放主体，也会对环境产生一定的影响。近年来，许多研究人员分析了人工智能训练模型，以估计训练它们所需的能源成本(以千瓦为单位)。通过将这种能源消耗转换为近似的碳排放和电力成本，研究者估计建成一个人工智能环保系统 AECOS 会产生约 30 万公斤的二氧化碳排放量，相当于乘坐飞机从纽约到北京往返 125 个来回航班所产生的碳排放量，而这需要同时种植 85-319 颗树才能抵消碳排放”。

在完成阅读任务后，要求被试报告 (1)“你认为人工智能环保系统 AECOS 的环境友好程度有多高”(1 = 完全没有, 7 = 非常高)以及(2) “人工智能环保系统 AECOS 对环境可能造成的破坏程度有多大” (1 = 完全没有, 7 = 非常大)作为操纵检查。接着，测量被试对人工智能环保系统 AECOS 的接受程度。采用与研究 1 相同的测量工具，请被试想象自己是某社区的管理者，并报告使用人工智能环保系统 AECOS 的意愿以及对人工智能环保系统 AECOS 的态度。同时，采用 Złotowski 等人(2017)编制的人工智能研究支持度的条目对被试进行测量，包括 “在多大程度上,你支持人工智能研究?”、“在多大程度上,你支持将纳税人的钱用于人工智能研究?”以及“在多大程度上,你支持国家划拨经费投入人工智能研究?”，题目均采用李克特 7 点计分 (1 = 强烈反对，7 = 强烈支持, α = 0.81)。

3.1.3 结果和讨论

操纵检验。操纵检查结果表明，低可持续性组 (M = 4.18, SD = 1.73)认为人工智能 AECOS 的环境友好程度显著低于控制组(M =5.67, SD = 0.96), $F(1, 237) = 67.78$, $p < 0.001$, $\eta^2 = 0.22$。同时，低可持续性组(M = 4.85, SD = 1.65)对环境破坏程度的评分显著高于控制组(M = 2.96, SD = 1.43), $F(1, 237) = 89.40$, $p < 0.001$, $\eta^2 = 0.27$，由此表明实验操纵有效。

主效应分析。对人工智能系统的接受程度进行方差分析检验(低可持续性=1, 控制组=0)，结果发现，低可持续性组(M = 4.62, SD = 1.56)的积极态度词评价得分显著低于控制组(M =5.79, SD = 0.60), $F(1, 237) = 57.96$, $p < 0.001$, $\eta^2 = 0.20$。同时，低可持续性组(M = 3.60, SD = 1.22)的消极态度词评价得分显著高于控制组(M = 2.41, SD = 1.22), $F(1, 237) = 47.20$, $p < 0.001$, $\eta^2 = 0.17$。并且，低可持续性组(M = 4.11, SD = 1.82)报告的使用意愿显著低于控制组

[3] 2023-07-18 取自 http://en.carbonstop.net/carbon_calculator_sch/

(M = 5.78, SD = 0.91), F(1, 237) = 81.33, p < 0.001, η^2 = 0.26。在控制年龄、性别、学历和职业后，低可持续性组的使用意愿依然显著低于控制组，F(1, 233) = 82.73, p < 0.001, η^2 = 0.26。此外，相比于控制组(M = 5.76, SD = 0.66)，低可持续性组(M = 5.32, SD = 1.25)对国家开展人工智能研究的支持度也更低, F(1, 237) = 11.26, p < 0.001, η^2 = 0.05。同样地，在控制年龄、性别、学历和职业后，两组被试在研究支持度的评分依然存在显著差异，F(1, 233) = 10.52, p = 0.001, η^2 = 0.04。综上，研究 1a 结果支持了研究假设 H1。

3.2 研究 1b

研究 1a 初步验证了可持续性对人工智能接受度的影响，但在该研究中人工智能环保系统态度测量条目是由 ChatGPT 3.5 生成的，其中积极词语的数量显著多于消极词语，因此这可能导致研究结果存在积极偏差。为了进一步验证研究假设 H1，研究 1b 将采用新的测量工具来探究人们对人工智能环保系统的态度。

3.2.1 被试

根据 G*power 的计算结果，至少需要 172 名被试才能保证研究达到中等效应量(f = 0.25, power = 0.90)。通过 Credamo 平台招募到 202 名被试参与本研究，删除 2 名未通过注意力检查的被试，剩下有效被试 200 名(其中男性 52 名, 女性 148 名; 18~62 岁; M_{age} = 30.95, SD = 7.38)。

3.2.2 研究过程

在知情同意后，被试随机进入到低可持续性人工智能组(n = 100)或控制组(n = 100)条件，并完成与研究 1a 相同的可持续性感知操纵任务和操纵检查测量。随后，采用 Stein 等人(2024)编制的人工智能态度问卷测量被试对人工智能环保系统 AECOS 的态度，并使用与研究 1a 相同的问题测量被试的使用意愿。Stein 等人(2024)编制的问卷共有 12 个条目，从认知、情感和行为三个维度测量了人们对人工智能的态度(α = 0.87)。为使其能够用于测量被试环保类型的人工智能的态度，研究 1b 对 12 个条目进行相应的修改，例如“人工智能环保系统 AECOS 会让社区变得更好”，“我对人工智能环保系统 AECOS 持强烈消极的态度”等(1=非常不同意, 5=非常同意)。在对第 2、4、7、8、10 题进行反向计分后计算所有条目的平均分，被试的得分越高，表明对人工智能环保系统 AECOS 的态度越积极。

同时，为了避免研究 1a 中积极态度词语和消极态度词语数量不均衡而导致的偏差，研究 1b 采用与研究 1a 相同的生成命令，在中文人工智能系统文心一言中生成了 14 个态度测量词，分为积极态度(4 个, α = 0.94)、中立态度(4 个, α = 0.32)、消极态度(4 个, α = 0.92)和矛盾态度(4 个, α = 0.32)不同的类别，具体包括“乐观的”、 “期待的”、“支持的”、“热情的”、“中立的”、“谨慎的”、“平淡的”、“客观的”、“担忧的”、“怀疑的”、“抵触的”、“消极的”、“矛盾的”和“心情复杂的”。在测量中，采取李克特 7 点计分，被试报告自己对各个条目的赞同程度(1 = 非常不同意, 7 = 非常同意)。

3.2.3 结果与分析

操纵检查。单因素方差分析结果表明(低可持续性=1, 控制组=0)，控制组的被试(M = 5.98, SD = 0.78)对人工智能环保系统 AECOS 的环境友好程度感知显著高于低可持续性组(M = 4.13, SD = 1.72)，$F(1, 198) = 94.46$，$p < 0.001$，$\eta^2 = 0.33$，而控制组(M = 2.69, SD = 1.43)对人工智能环保系统 AECOS 的环境破坏程度的感知显著低于低可持续性组(M =5.08, SD = 1.43)，$F(1, 198) = 140.41$，$p < 0.001$，$\eta^2 = 0.42$。这表明，对可持续性的实验操纵是有效的。

主效应分析。以使用意愿作为因变量进行单因素方差分析，发现控制组被试(M = 5.93, SD = 0.89)对人工智能参与环境决策的接受度显著高于低可持续性组(M = 4.29, SD = 1.71)，$F(1, 198) = 73.33$，$p < 0.001$，$\eta^2 = 0.27$。在控制年龄、性别、学历和职业后，两组被试在人工智能使用意愿的得分依然存在显著差异，$F(1, 194) = 57.37, p < 0.001, \eta^2 = 0.23$。

对人工智能环保系统 AECOS 的态度差异进行分析，发现控制组被试(M = 4.02, SD = 0.37)对人工智能环保系统 AECOS 的态度评分显著高于低可持续性组(M = 3.31, SD = 0.80)，$F(1, 198) = 66.07$，$p < 0.001$，$\eta^2 = 0.25$。在控制年龄、性别、学历和职业后，两组被试在人工智能态度的评分依然存在显著差异，$F(1, 194) = 55.68, p < 0.001, \eta^2 = 0.22$。这表明，低可持续性降低人们对人工智能的积极态度评价。

通过对文心一言生成的态度词语评分的差异性进行分析，发现控制组被试(M = 5.82, SD = 0.69)对人工智能环保系统 AECOS 的积极态度评分显著高于低可持续性组(M = 4.43, SD = 1.55)，$F(1, 198) = 66.64$，$p < 0.001$，$\eta^2 = 0.25$，在控制年龄、性别、学历和职业后，两组被试在人工智能积极态度的评分依然存在显著差异，$F(1, 194) = 52.79, p < 0.001, \eta^2 = 0.21$。同时，控制组被试($M$ = 2.43, SD = 0.90)对人工智能环保系统 AECOS 的消极态度评分显著低于低可持续性组(M = 3.90, SD = 1.54)，$F(1, 198) = 68.05$，$p < 0.001$，$\eta^2 = 0.26$，而在控制年龄、性别、学历和职业后，两组被试在人工智能消极态度的评分依然存在显著差异，$F(1, 194) = 56.51, p < 0.001, \eta^2 = 0.23$。所以，研究 1b 再次表明，可持续性影响人们对人工智能的使用意愿和态度：可持续性程度越低，人们对人工智能环保系统参与环境决策的接受度和积极态度就越低，而对人工智能环保系统的消极态度越强烈。

4 研究 2

研究 1 通过操纵被试对可持续性的感知(有 vs.无)，在 2 个子研究中验证了人工智能的可持续性和接受意愿的关系，并发现低可持续会降低人们对人工智能参与环境决策的支持度和态度评价。然而，研究 1 中实验组和控制组的被试所阅读的材料长度并非完全一致，这可能在一定程度上影响了实验结果。换言之，前述研究未在可持续性的高低水平上进行操纵和对比，因此无法很好地说明可持续的属性对环保人工智能接受度的影响。因此，研究 2 将通过向被试呈现信息以启动高可持续性(“绿色 AI”)和低可持续性(“红色 AI”)的感知，并探索其中的中介机制。

4.1 被试

通过 Credamo 平台招募到 300 名被试参与本研究，7 名被试因未通过注意力筛查而被剔除，剩下有效被试 293 名(男性 90 名, 女性 203 名, 18~64 岁; *M*age = 30.26 岁，*SD* = 9.30)。

4.2 研究程序

在经过知情同意程序后，被试首先报告了个人的基本信息，然后随机进入到低可持续性组(红色 AI, $n = 147$)或高可持续性组(绿色 AI, $n = 146$)。其中，红色 AI 组阅读与研究 2 相同的低可持续相关信息；绿色 AI 组首先阅读一段内容以了解人工智能在环保领域的应用："近年来，越来越多地组织开始利用自主人工智能(AI)来帮助解决环境问题。其中，人工智能环保系统 AECOS 就是一种专门解决环保问题的智能系统，它使用算法来推断特定问题的规则，并根据社区各个垃圾处理站的运营情况、社区协管员的工作表现和时间安排进行自动评分，以保证做出的最佳决策"。随后，继续向被试呈现以下信息(参考 Verdecchia et al., 2023): "但是，人工智能本身也是重要的碳排放主体，也会对环境产生一定的影响。因此，为了更高效地利用大数据资源，人工智能环保系统 AECOS 不仅功能强大，更在系统设计和运行中充分纳入了环保和生态指标，所以它比同类型的智能环保系统节约 20%的计算成本，产生能源消耗和碳排放量也低 20% "。

阅读信息后，被试完成与研究 1a 相同的操纵检查，报告人工智能环保系统 AECOS 的环境友好程度和对环境可能造成的破坏程度。随后，采用与研究 1a 相同的题目，测量被试使用人工智能环保系统 AECOS 的意愿以及对人工智能研究的支持度($\alpha = 0.86$)。最后，测量被试对人工智能环保系统 AECOS 的道德能动性感知(Banks, 2019)，包含道德(morality)和依从(dependency)两个维度，例如"人工智能环保系统 AECOS 对于对和错是有感觉的"、"人工智能环保系统 AECOS 的行为符合道德规范" 以及"人工智能环保系统 AECOS 只能做人类告诉它的事情"等共 10 个条目(1 = 非常不同意, 7 = 非常同意)。其中，道德维度有 6 个条目 ($\alpha = 0.91$)，测量了被试对人工智能的理性和道德直觉的感知；依从维度有 4 个条目 ($\alpha = 0.86$)，所有题项在语义上都与人工智能缺乏自主性和意图性相关，以表明它们是被动的，并且依赖于编程或其他人的控制。在完成所有题目后，被试均获得一定的报酬。

4.3 结果与讨论

操纵检验。以人工智能可持续性作为自变量，环境友好程度和破坏程度分别作为因变量进行方差分析，结果发现，红色 AI 组($M = 3.94$, $SD = 1.71$)的环境友好程度显著低于绿色 AI 组($M = 5.68$, $SD = 0.81$), $F(1, 291) = 124.04$, $p < 0.001$, $\eta^2 = 0.30$；红色 AI 组($M = 4.99$ $SD = 1.48$)的环境破坏程度显著高于绿色 AI 组($M = 2.99$, $SD = 1.31$), $F(1, 291) =150.12$, $p < 0.001$, $\eta^2 = 0.34$。由此表明，实验操纵是有效的。

主效应分析。与假设 H2 一致，红色 AI 组($M = 4.50$, $SD = 0.11$)使用人工智能环保系统 AECOS 的意愿显著低于绿色 AI 组($M = 5.89$, $SD = 0.11$)，$F(1, 291) = 82.51$, $p < 0.001$, $\eta^2 = 0.22$。在控制年龄、性别、学历和职业后，两组的使用意愿也依然存在显著差异，$F(1, 287) = 78.91$, $p < 0.001$, $\eta^2 = 0.22$。并且，红色 AI 组($M = 5.15$, $SD = 1.17$)被试对于国家开展

人工智能研究的支持度也显著低于绿色 AI 组(M = 5.55, SD = 0.86)，$F(1, 291) = 11.61, p = 0.001, \eta^2 = 0.04$。在控制年龄、性别、学历和职业后，两组被试的支持度也依然存在显著差异，$F(1, 287) = 10.82, p = 0.001, \eta^2 = 0.04$。由此，研究 2 重复了研究 1 的结果，并再次支持了研究假设 H2。

分别以道德和依从作为因变量进行方差分析检验，与预期结果一致，红色 AI 组被试(M = 4.15, SD = 1.36)对人工智能的道德评价显著低于绿色 AI 组(M = 4.58, SD = 1.33)，$F(1, 291) = 7.35, p = 0.007, \eta^2 = 0.03$；但红色 AI 组被试($M$ =5.41, SD = 1.04)对人工智能的依从程度评价高于绿色 AI 组(M = 5.07, SD = 1.36)，$F(1, 291) = 6.10, p = 0.014, \eta^2 = 0.02$。即，相比于绿色 AI 组的被试，红色 AI 组被试认为人工智能环保系统 AECOS 的能动性更低。这表明，低可持续性会降低人们对人工智能的道德感知和能动感知。

中介效应分析。以道德和依从作为中介变量，人工智能的可持续性作为自变量，使用意愿作为因变量，在 PROCESS 4.0 插件中选择模型 4 进行多重中介效应检验。结果发现，可持续性对道德预测作用显著，b*=－0.31, SE = 0.16，95% CI % [－0.74, － 0.12]，道德对使用意愿的预测作用显著， b*= 0.32, SE = 0.06，95% CI % [0.23, 0.46]，此时道德的中介效应值 B = －0.15，SE = 0.07，95% CI [－0.29, －0.03] 不包含 0。尽管依从的间接效应显著，B = 0.05，SE = 0.03，95% CI [0.005, 0.113]不包含 0，但此时可持续性对使用意愿的直接效应值为－1.29(SE = 0.15，95% CI －1.58, －1.00])且符号与间接效应符号相反，这说明依从在人工智能的可持续性和使用意愿之间起到抑制作用而非中介作用(MacKinnon et al., 2000)，具体路径系数如图 1 所示。所以，道德是解释可持续性与接受程度之间的中介机制，支持了研究假设 H2a。

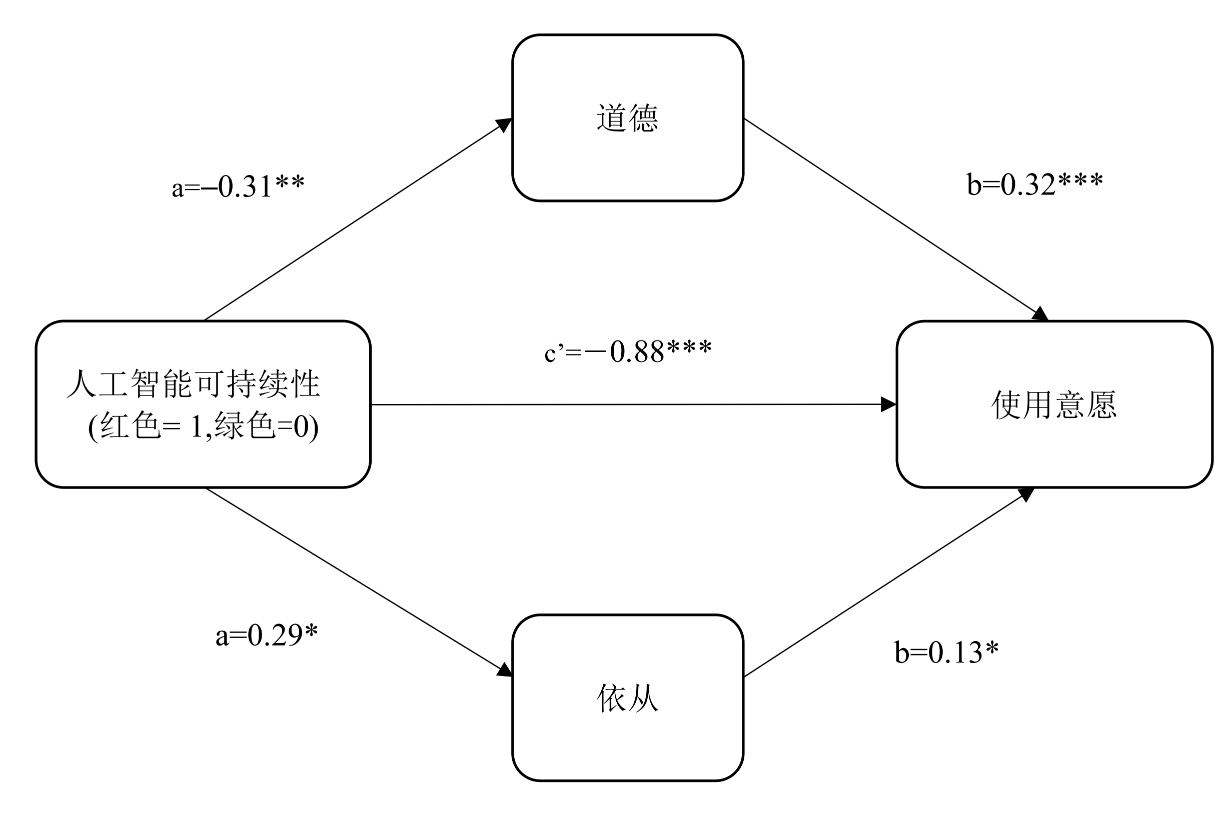


图 **1** 研究 2 中道德和依从的中介作用

注：$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; 图中系数值为标准回归系数

以对开展人工智能研究的支持度为因变量，对道德和依从的中介作用进行检验。再次发现，道德的中介效应显著，*B* = − 0.12，*SE* = 0.05，95% CI [−0.24, −0.03] 不包含 0，表明道德在人工智能可持续性与研究支持度之间也起到显著的中介作用。同样的，尽管依从的间接效应显著值 *B* = 0.04，*SE* = 0.02，95% CI [0.01, 0.10]不包含 0，但此时可持续性对研究支持度的直接效应值为−0.34(*SE* = 0.12，95% CI [−0.56, −0.11])且符号与间接效应符号相反，这说明依从在人工智能的可持续性和研究支持度之间起到抑制作用而非中介作用 (MacKinnon et al., 2000)，具体路径系数如图 2 所示。所以，道德是解释可持续性与接受度之间的中介机制，再次支持了研究假设 H2a。

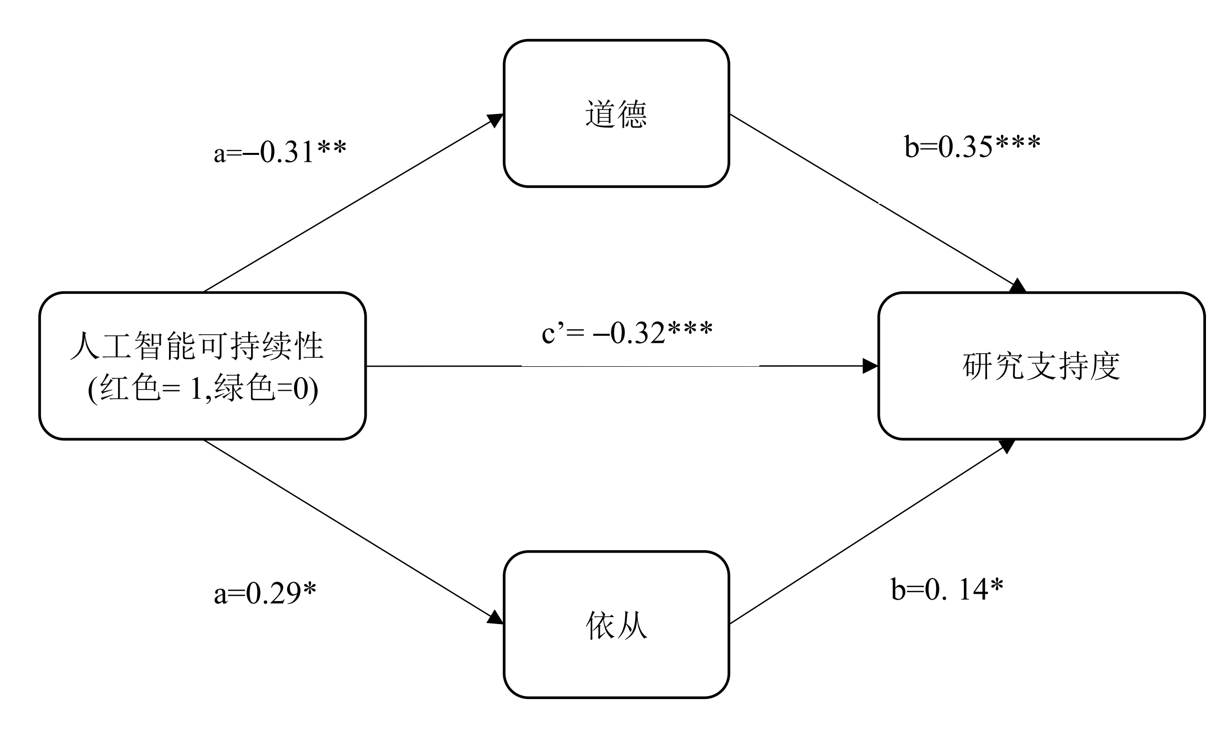


图 **2** 研究 2 中道德和依从的中介作用

注：$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; 图中系数值为标准回归系数

综上所述，研究 2 通过实验操纵了人工智能的(高 vs. 低)可持续性，验证了可持续性对环境领域人工智能接受度的影响，即低可持续性会降低人们对人工智能的使用意愿和研究支持度。更重要的是，研究 2 探索了影响二者关系的中介机制，并发现相比于依从 (即低能动性)，对人工智能的道德感知才是解释可持续性和人工智能接受度之间关系的中介变量。

5 研究 3

通过前 3 个研究可知，低可持续的人工智能会降低人们对其的道德感知，进而降低对人工智能的接受意愿。然而，这一影响过程可能受到个体差异的影响，例如个体的亲环境

态度越强烈，低可持续性对接受意愿的影响就大。因此，研究 3 将通过实验操纵人工智能的可持续性，以检验亲环境态度的调节作用，并再次重复研究 1 和研究 2 的结果。

5.1 被试

根据前述研究，在保证被试量达到 G*Power 3.1 计算出的最小样本量为 172 的基础上($f = 0.25$, power = 0.90)，通过 Credamo 随机招募到 306 名被试。由于 9 名被试未通过注意力筛查，2 名被试年龄小于 18 岁而被剔除，最终有效被试为 295 名(男性 132 名, 女性 163 名; 18~67 岁, $M_{age} = 31.43$ 岁，$SD = 9.65$)。

5.2 研究程序

在知情同意程序后，被试首先报告个人基本信息，然后被随机分配进入到人工智能可持续性低(红色 AI, $n = 148$)或可持续性高组(绿色 AI, $n = 147$)的实验条件下，具体实验操纵和测量条目与研究 2 相同。在完成操纵检查后，被试报告对人工智能系统的接受程度，包括使用人工智能环保系统 AECOS 的可能性和对国家开展人工智能研究的态度(具体条目与研究 2 相同, $\alpha = 0.91$)。随后，采用与研究 2 相同的条目，测量被试对人工智能环保系统 AECOS 的道德($\alpha = 0.94$)和依从($\alpha = 0.91$)程度的感知。接着，采用由 Dunlap 等人(2000)编制的新生态范式量表(New Ecological Paradigm Scale)测量被试的亲环境态度，共 15 题，例如 “当前人口数量即将接近地球有效的承载力”、“人类面临的所谓生态危机被极度地夸大了”(1 = 非常不同意, 5 = 非常同意, $\alpha = 0.84$)，并且在数据分析时对偶数题项进行反向计分。在完成所有题目后，被试获得一定的报酬作为感谢。

5.3 结果与讨论

操纵检查。以人工智能可持续性作为自变量(低可持续性=1, 高可持续性=0)，环境友好程度和破坏程度分别作为因变量的方差分析，结果表明红色 AI 组($M = 3.86$, $SD = 1.81$)对环境友好程度的评分显著低于绿色 AI 组($M = 5.82$, $SD = 0.86$), $F(1, 293) = 140.90$, $p < 0.001$, $\eta^2 = 0.33$；而且，红色 AI 组($M = 5.00$, $SD = 1.58$)对环境破坏程度的评分显著高于绿色 AI 组($M = 3.15$, $SD = 1.51$), $F(1, 293) = 106.19$, $p < 0.001$, $\eta^2 = 0.27$。因此，实验操纵是有效的。

主效应分析。对使用意愿和研究支持度分别进行方差分析检验，结果显示，红色 AI 组($M = 4.19$, $SD = 1.85$)使用人工智能环保系统 AECOS 的意愿显著低于绿色 AI 组($M = 5.82$, $SD = 0.88$)，$F(1, 293) = 93.06$, $p < 0.001$, $\eta^2 = 0.24$。在控制年龄、性别、学历和职业后，两组的使用意愿也依然存在显著差异，$F(1, 289) = 97.32$, $p < 0.001$, $\eta^2 = 0.25$。并且，相比于绿色 IA 组($M = 5.63$, $SD = 0.99$)，红色 AI 组($M = 4.78$, $SD = 1.56$)被试对国家开展人工智能研究的支持度也显著更低，$F(1, 293) = 31.15$, $p < 0.001$, $\eta^2 = 0.10$。同样地，在控制年龄、性别、学历和职业后，两组的支持度也依然存在显著差异，$F(1, 289) = 30.72$, $p < 0.001$, $\eta^2 = 0.10$。由此，研究 3 重复了研究 1 和研究 2 的结果，并再次验证了研究假设 H1。

分别以道德和依从作为因变量进行方差分析检验，与预期结果一致，红色 AI 组被试($M = 4.08$, $SD = 1.64$)对人工智能的道德评价显著低于绿色 AI 组($M = 4.56$, $SD = 1.41$)，$F(1,$

293) = 8.15, p = 0.005, η^2 = 0.03；并且，红色 AI 组被试(M = 5.33, SD = 1.27)对人工智能的依从程度评价显著高于绿色 AI 组(M = 4.88, SD = 1.55)，F(1, 293) = 7.51, p = 0.03, η^2 = 0.02。这表明，低可持续性会降低人们对人工智能的道德感知，而增强对人工智能的依从程度感知。

中介效应分析。将使用意愿作为因变量，道德和依从作为中介变量，在 PROCESS 4.0 插件中选择模型 4 进行多重中介效应检。结果显示，可持续性对道德的预测作用显著, b*= －0.33, SE = 0.18，95% CI % [－0.86, － 0.16]，道德对使用意愿的预测作用显著， b*= 0.43, SE = 0.05，95% CI % [0.37, 0.55]，此时道德的间接效应显著，B =－0.25，SE = 0.10，95% CI [－0.45, －0.07] 不包含 0，表明道德在可持续性和使用意愿之间起到中介作用，支持了研究假设 H2a。尽管依从的间接效应值 B = 0.05，SE = 0.03，95% CI [0.005, 0.2] 不包含 0，但此时可持续性对使用意愿影响的直接效应值为－1.43(SE = 0.15，95% CI － 1.73, －1.14])且符号与间接效应符号相反，这说明依从在人工智能的可持续性和使用意愿之间起到抑制作用而非中介作用(MacKinnon et al., 2000)，研究假设 H2b 未被验证。所以，道德是解释可持续性与接受意愿之间关系的中介机制，研究 3 重复了研究 2 的结果，支持了研究假设 H2a。

同样地，以研究支持度作为因变量，对道德和依从的中介作用进行检验。结果发现，可持续性对道德的预测作用显著, b*=－0.33, SE = 0.18，95% CI % [－0.86, － 0.16]，道德对研究支持度的预测作用显著， b*= 0.41, SE = 0.05，95% CI % [0.32, 0.50]，此时道德的间接效应显著 B =－0.21，SE = 0.08，95% CI [－0.38, －0.06] 不包含 0。尽管此时依从的间接效应 B = 0.09，SE = 0.04，95% CI [0.02, 0.16])，但间接效应值的符号与直接效应值(B =－0.73，SE = 0.14，95% CI [－1.00, －0.46])符号相反，说明依从在其中起到抑制而非中介作用(MacKinnon et al., 2000)，再次验证了研究假设 H2a。因此，研究 3 与研究 2 的结果一致，再次表明，相比于对人类或程序的依从，人们更重视人工智能的道德，因此道德是解释可持续性和人工智能接受意愿之间关系的心理因素。

调节效应分析。以使用意愿作为结果变量，道德作为中介变量，亲环境态度作为调节变量，采用 PROCESS4.0 中的模型 8 进行有调节的中介效应检验。结果显示，可持续性与亲环境行为态度的交互作用对使用意愿的影响显著，B =－1.63, SE = 0.26, 95% CI [－2.13, －1.12]，支持了研究假设 H3a。同时，可持续性与亲环境态度的交互作用显著负向预测道德感知，b = －0.95, SE = 0.31, 95% CI [－1.57, －0.34]，表明亲环境态度调节可持续性和道德感知之间的关系。并且，有调节的中介效应值 index 为－0.36, SE = 0.14, 95% CI [－0.65, －0.10]，表明亲环境态度对可持续性通过道德感知影响使用意愿的路径具有显著的调节作用，验证了研究假设 H3b。

进一步地进行简单效应检验，如图 **3** 所示，亲环境态度越强烈，低可持续性对人工智能的道德感知负向影响就越大， b = －1.12, SE = 0.24, 95% CI [－1.59, －0.64]；随着亲环

境态度的减弱，低可持续性降低对道德感知的负向影响随之消失，$b = -0.07, SE = 0.24$, 95% CI [−0.55, 0.41]，这一结果符合与研究假设 H3b 一致。

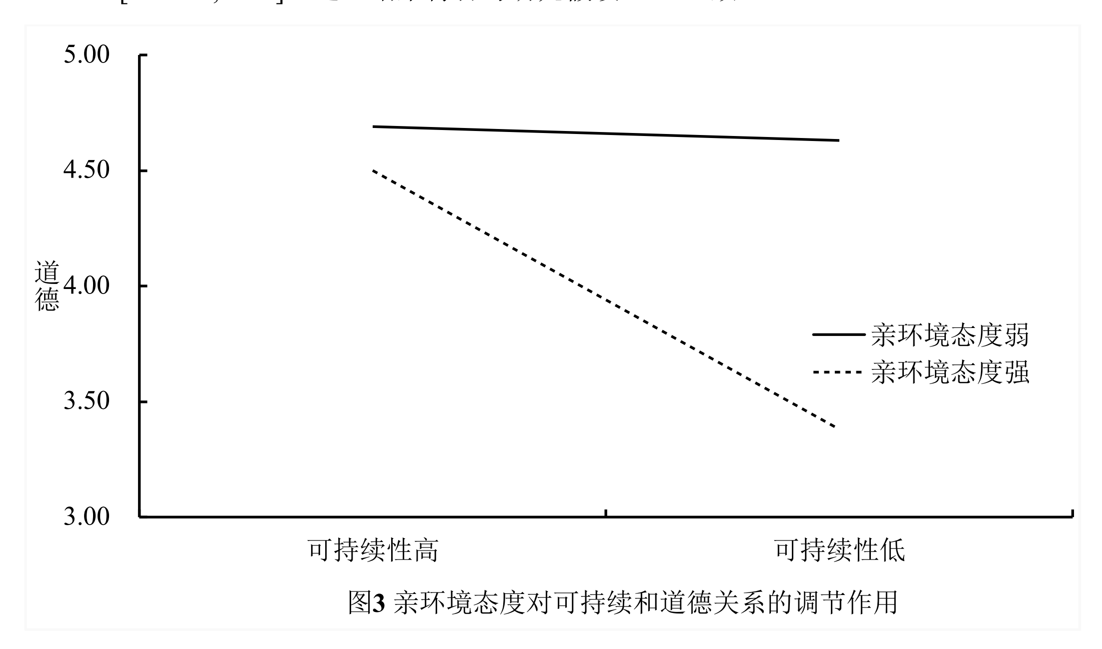


图3 亲环境态度对可持续和道德关系的调节作用

并且，如图 4 所示，对于亲环境态度强烈的人而言，低可持续性对人工智能使用意愿的负向影响更大，$b = -2.37, SE = 0.20$, 95% CI [−2.77, −1.98]；对于亲环境态度较弱的个体而言，低可持续性对人工智能使用意愿的负向影响更小，$b = -1.12, SE = 0.24$, 95% CI [−1.59, −0.64]，从而验证了研究假设 H3a 和 H3b。

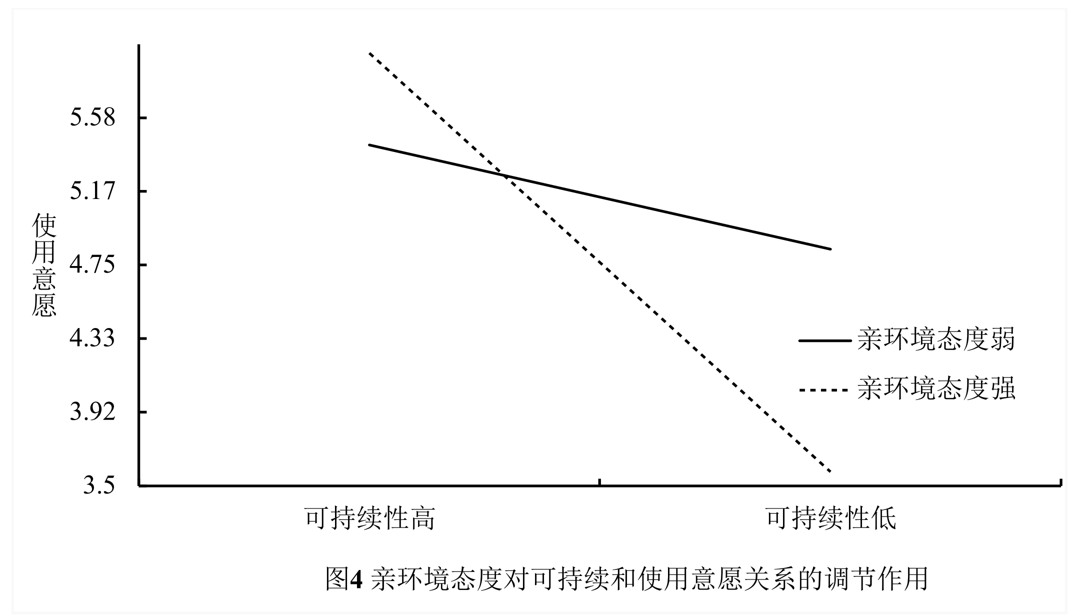


图4 亲环境态度对可持续和使用意愿关系的调节作用

以研究支持度作为结果变量，道德作为中介变量，亲环境态度作为调节变量，再次采用 PROCESS4.0 中的模型 8 进行有调节的中介效应检验。结果显示，可持续性与亲环境行为态度的交互作用对研究支持度的影响显著，$B = -0.79, SE = 0.25$, 95% CI [−1.29, −0.29]，再次支持了研究假设 H3a。同时，可持续性与亲环境态度的交互作用显著负向预测道德感

知，*b* = －0.95, *SE* = 0.31, 95% CI [－1.57, －0.34]，表明亲环境态度调节可持续性和道德感知之间的关系。并且，有调节的中介效应值 index 为－0.34, *SE* = 0.14, 95% CI [－0.64, －0.09]，表明亲环境态度对可持续性通过道德感知影响研究支持度的路径具有显著的调节作用，再次验证了研究假设 H3b。

同样地，如图 **5** 所示，对于亲环境态度强烈的人而言，低可持续性对人工智能研究支持度的负向影响更大，*b* = －1.07, *SE* = 0.20, 95% CI [－1.46, －0.68]；对于亲环境态度较弱的个体而言，低可持续性对人工智能研究支持度的负向影响越小，*b* =－0.21, *SE* = 0.19, 95% CI [－0.59, －0.68]，从而验证了研究假设 H3a 和 H3b。

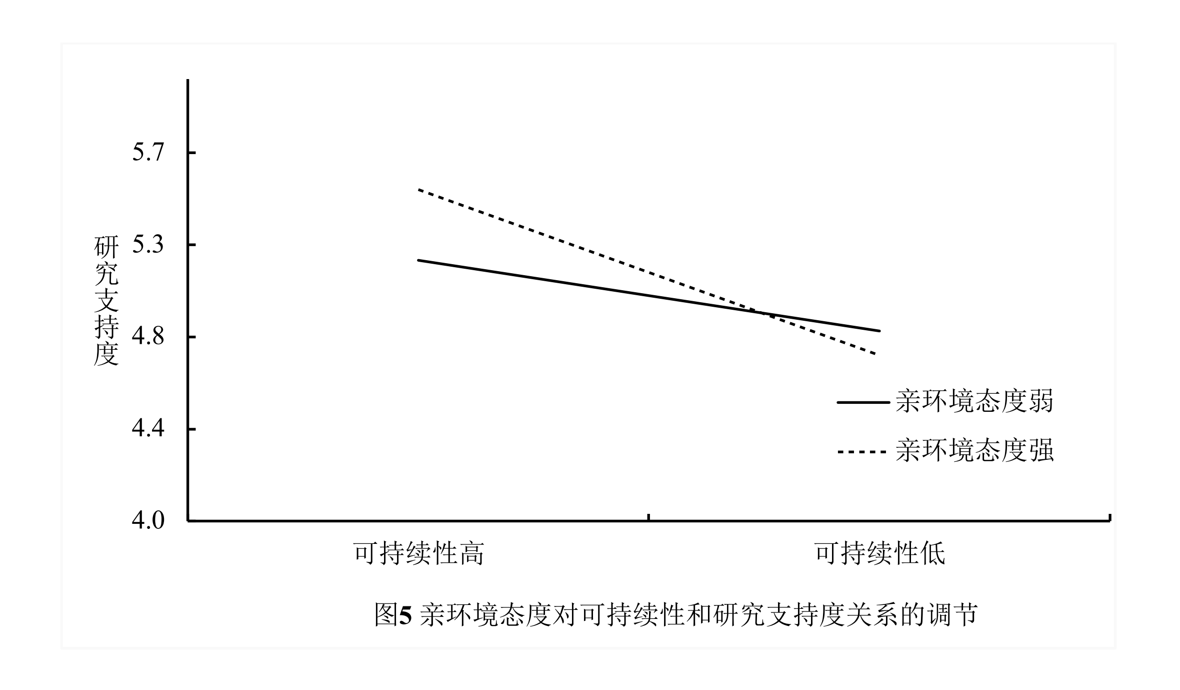


图**5** 亲环境态度对可持续性和研究支持度关系的调节

综上，研究 3 再次表明低可持续性降低人们对人工智能的接受意愿，因为人们认为低可持续性的人工智能道德水平更低，而这一影响在亲环境态度强烈的群体中表现更为突出。

6 总讨论

基于技术接受与使用统一模型，本研究考察了可持续性对人工智能接受意愿的影响及机制。通过 5 个递进的子研究，发现人们对于将人工智能应用于环境决策总体持以积极态度，且有较高的使用意愿(预研究)。但是，低可持续性会降低人们对人工智能的使用意愿和研究支持度 (研究 1a 和研究 1b、研究 2 和研究 3)。这是因为，低可持续性导致人们对人工智能的道德感评价降低 (研究 2 和研究 3)。此外，亲环境态度越强烈，人们不仅对低可持续人工智能的道德评价越低，而且对低可持续性人工智能的接受意愿也越低(研究 3)。

6.1 人工智能和可持续发展

从技术层面来讲，人工智能对于社会可持续发展有着重要意义(Nishant et al., 2020; Vinuesa et al., 2020)，而本研究发现人类在心理上也愿意使用人工智能来解决环保问题。并且，人们对人工智能还有道德层面的要求，即遵守洁净和关爱的道德规范。所以，当意识到人工智能存在高耗能和高碳排放的问题后，人们使用人工智能的意愿降低，对国家开展人工智能研究的支持度也显著降低。这一结果拓展了人工智能和可持续发展关系的研究视角，从心理学角度补充了人类对人工智能参与环保决策的态度和反应 (Nishant et al., 2020)，表明在人们在心理上也认为人工智能可以作为人类同盟共同进行管理决策 (如 Haesevotes et al., 2021)。

以往研究发现，可持续性对智能家居的影响并不显著(Ahn et al., 2016; Baudier et al., 2020)。但是，Gansser 和 Reich(2021)在探讨影响人们使用人工智能相关产品的行为意向和使用行为的因素时，发现人工智能产品的可持续性(如, 废物管理、节约资源和能源消耗等)对使用意愿和使用行为的影响虽然微小但显著。研究 1 到研究 3 通过不同方式的实验操纵验证了低可持续性对人工智能使用意愿存在消极且相对稳定的影响，支持并扩展了 Gansser 和 Reich(2021)的研究。并且，与 Schwartz 等人(2020)和 van Wynsberghe(2021)的观点一致，本研究表明，实现可持发展发展不仅需要人工智能“向善”(促进可持续发展)，同时也需要人工智能“为善”(具有高可持续性的属性)。

6.2 道德能动性的中介作用

通过研究 2 和研究 3，我们发现，相比于依从(H2b)，人们对人工智能的道德感知(H2a)更可能决定对其的接受意愿。因此，与以往研究者的观点一致，人工智能应当与人类的价值和道德规范对齐(Djeffal et al., 2022)，比如要符合环境友好、资源节约的要求(Gansser & Reich, 2021; Siala & Wang, 2022)。低可持续性的人工智能违反这一道德要求，可能导致人们对其产生更低的道德评价，从而降低个体的使用意愿和开展研究的支持意愿。

虽然人工智能是否拥有道德能动性还存在较多争议，但是人类内隐地将人工智能视为具有道德地位的道德实体，并且试图用人类的道德价值对人工智能(发出或接受)的行为进行感知和判断(喻丰, 许丽颖, 2018; Banks, 2019)。研究 2 和研究 3 中，感知到人工智能是低可持续性的被试对人工智能的道德评估显著更低，而且只有道德(相比于高依从, 即低能动)在可持续性和接受度之间起到中介作用。这支持了 Bank(2019)的研究，即感知到的道德水平越高，人们与人工智能接触的意愿越强烈，而依从与人工智能的接触意愿之间不相关或可能存在负相关关系(Haikonen, 2007)。这一结果也符合社会感知理论的基本假设，道德在社会认知中发挥着主要作用 (Ray et al., 2021)。同时，保护环境是人类普遍应当遵守的神圣价值规范(Tetlock, 2003)，低可持续性的人工智能因违反这一道德规范会引发人们强烈的道德愤慨，从而导致人们通过拒绝使用低可持续的人工智能来维护保护环境的价值规范(Tetlock, 2002)。所以，本研究发现道德在可持续性和人工智能接受度之间起到中介作用。

同时，依从(低能动)可能并非是解释可持续性和接受意愿关系的潜在机制，这一结果从心理学视角补充了人工智能道德能动性的实证研究。尽管人类期望人工智能具有道德行为能力且能够承担道德责任，但人工智能本身并不会或不能做出(不)道德行为(Anderson & Anderson, 2007)，而是依赖于人类对其的训练和程序设置(Moor, 2006)。低可持续性的人工智能通常是高耗能和高碳排放的(Schwartz et al., 2020; van Wynsberghe, 2021)，这意味着低可持续性的人工智能需要更多的能源和资源来保证其运行工作，反映出其对外部输入的高度依赖，而这可能被会感知为依从性较高。同时，低可持续性意味着人工智能系统的运行效率较低，可能需要频繁的维护和人类干预，这进一步强化了人们对其依从性高、自主性低的感知。而且，本研究中的人工智能是在社区环境管理情境中的智能系统，其本质是弱人工智能，而非能够沟通、能以自主的方式行事且具有判断、反思和承诺能力的主体(Constantinescu et al., 2022; Swanepoel, 2021)。所以，在研究 2 和研究 3 中，我们发现低可持续性的人工智能被感知为依从性更高(即自主性更低)，这种依从性虽然间接提高了使用意愿，但低可持续性对使用意愿的直接负面影响更强，从而形成了抑制效应。这说明，尽管人们期望人工智能具有道德能动性，但当低可持续的人工智能显示出高度的依从性、低道德水平时，人们可能认为人工智能不具备能动性且其带来的伤害是程序操控导致的(Wilson et al., 2022)，从而减少人们对其道德责任的归因 (例如, Maninger & Shank, 2022; Monroe et al., 2014) ，进而影响其接受意愿。

6.3 亲环境态度的调节作用

研究 3 发现亲环境态度分别调节了可持续性和道德的关系，以及可持续性和接受意愿之间的关系。具体地，对环境或环境问题关怀程度越高，个体不仅对低可持续性人工智能的道德感知越低，对其的接受度也越低。由于暂未有研究从心理学角度考察人类对于环保领域人工智能有何行为反应(Nishant et al., 2020)，因此本研究为了解人类对人工智能碳排放问题的心理和行为反应提供了初步的实证资料。

环境问题的本质是道德问题，而个体对环境问题的态度和行为与其道德价值观和道德信念有关(Feinberg & Willer, 2013; Jia et al., 2017)，因此做出保护环境行为的人更可能被视为环保态度积极的且道德的(Braun Kohlova & Urban, 2020; Urban et al., 2023)。与以已有研究结果一致，在本研究中，阅读到人工智能产生高耗能和高碳放信息的被试对人工智能的道德评价也更低。同时，我们还发现，亲环境态度强烈的个体更可能拒绝使用低可持续性的人工智能，这一结果支持了以往研究结果(Averdung & Wagenfuehrer, 2011; Camilleri et al., 2019; De Canio, 2023) ，为亲环境态度和亲环境行为的积极关系提供了支持证据 (Wyss et al., 2022)。

6.4 研究意义和展望

理论上，本研究首先丰富了人工智能接受意愿前因的研究，检验和补充了可持续性对接受意愿的影响。尽管已有许多研究以技术接受模型或者其变体为理论基础探讨了人工智

能技术相关特征和用户个人特质对接受意愿的影响(Kelly et al., 2023)，但少有研究者关注人工智能的可持续性与人工智能接受意愿的关系。因此，本研究通过多个研究验证了可持续性同样是影响人工智能接受意愿的重要因素。其次，本研究从心理学的角度探究了人类对于人工智能参与环境决策的态度，丰富了人工智能和可持续发展的研究视角。虽然人工智能有助于推动实现可持续发展目标(Vinuesa et al., 2020)，然而鲜有研究者探讨人类对于人智能参与解决环保问题的心理和行为反应(Nishant et al., 2020)。本研究从环境心理学的角度，考察了社区环境治理情境下，高可持续性对人工智能接受意愿的积极作用和道德的中介作用。

本研究具有一定的现实意义：首先，为人工智能的设计和开发提供了实证依据。以往已有研究者指出，人工智能的研究在考虑算力的时候也应当考虑效能(efficiency)，本研究表明人类在考虑使用人工智能解决现实问题时，不仅强调计算力和有用性，同样看重人工智能的效能和可持续性(Schwartz et al., 2020; van Wynsberghe, 2021)，所以设计高可持续性的人工智能可以显著提高人类接受人工智能同伴的意愿。其次，为人工智能的社会治理提供心理学研究证据。当前世界范围内出现了多种人工智能治理准则，且强调人工智能应当与人类的道德准则对齐，本研究表明，当人工智能因破坏环境而违反人类的道德准则时，人类对其的道德评价和接受度都会随之降低，因此人工智能发展应当符合友好、资源节约的要求[2]。

本研究也存在一些不足：(1)尽管我们尝试对人工智能的可持续性做出概念界定，但本研究给出的定义可能依然是不够清晰和准确的; (2)预研究和研究 1 采用了由人工智能生成的态度评价词语作为测量工具，其信效度还需在未来研究中进行进一步的检验。并且，更高阶版本的人工智能所生成的态度形容测量词可能具有更高的信效度，而这有助于提高研究结果的准确性，因此未来研究可以使用更智能的 AI 系统来辅助人类开展研究；(3)预研究中使用的实验材料未考虑人工智能对该领域工作人员可能产生的职业威胁等现实问题。同时，我们也意识到研究 2 中的实验材料可能存在框架效应，低可持续实验条件下的被试在阅读实验材料时可能产生更多的消极情绪体验，而这也可能影响了实验结果。最后，未来研究可以考虑在人类和人工智能同时违反环保节约的规范情境下人们对不同主体的态度和行为反应，以提高研究的应用范围和现实意义。

7 结论

综上所述，本研究发现：(1)人们对于使用人工智能解决环保问题普遍持有积极态度; (2)低可持续降低人们环保人工智能的积极评价，而且对其参与环境决策的支持意愿和对国家开展人工智能研究的支持度也更低；(3)究其原因，低可持续性降低了人们对人工智能的道德评价，进而导致人们的使用意愿和研究支持度的降低；(4)对于环保态度强烈的群体而言，前述这一效应表现更为突出：亲环境态度越强烈，人们对低可持续性人工智能的道德感知越低，进而降低对人工智能的使用意愿和研究支持度。

参考文献

Ahn, M., Kang, J., & Hustvedt, G. (2016). A model of sustainable household technology acceptance. *International Journal of Consumer Studies, 40*, 83–91. https://doi.org/10.1111/ijcs.12217

Al-Sharafi, M. A., Al-Emran, M., Arpaci, I., Iahad, N. A., AlQudah, A. A., Iranmanesh, M., & Al-Qaysi, N. (2023). Generation Z use of artificial intelligence products and its impact on environmental sustainability: A cross-cultural comparison. *Computers in Human Behavior*, *143*, 107708. https://doi.org/10.1016/j.chb.2023.107708

Anderson, M., & Anderson, S.L. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, *28*(4), 15–26. https://doi.org/10.1609/aimag.v28i4.2065

Averdung, A., & Wagenfuehrer, D. (2011). Consumers' acceptance, adoption and behavioural intentions regarding environmentally sustainable innovations. *E3 Journal of Business Management and Economics*, *2*(3), 98–106.

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, *90*, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Baudier, P., Ammi, C., & Deboeuf-Rouchon, M. (2020). Smart home: Highly-educated students' acceptance. *Technological Forecasting and Social Change*, *153*, 119355. https://doi.org/10.1016/j.techfore.2018.06.043

Bigman Y. E., Wilson D., Arnestad M. N., Waytz A., Gray K. (2023). Algorithmic discrimination causes less moral outrage than human discrimination. *Journal of Experimental Psychology: General*, *152*(1), 4–27. https://doi.org/10.1037/xge0001250

Braun Kohlová, M., & Urban, J. (2020). Buy green, gain prestige and social status. *Journal of Environmental Psychology*, *69*, 101416. https://doi.org/10.1016/j.jenvp.2020.101416

Bretter, C., Unsworth, K. L., Kaptan, G., & Russell, S. V. (2023). It is just wrong: Moral foundations and food waste. *Journal of Environmental Psychology*, *88*, 102021. https://doi.org/10.1016/j.jenvp.2023.102021

Camilleri, A.R., Larrick, R., Hossain, S., & Patino-Echeverri, D. (2019). Consumers underestimate the emissions associated with food but are aided by labels. *Nature Climate Change, 9*, 53–58. https://doi.org/10.1038/s41558-018-0354-z

Chen, S.H., Qiu, H., Xiao, H., He, W., Mou, J., & Siponen, M.T. (2020). Consumption behavior of eco-friendly products and applications of ICT innovation. *Journal of Cleaner Production*, *287*,125436. https://doi.org/10.1016/j.jclepro.2020.125436

Constantinescu, M.V., Vică, C., Uszkai, R., & Voinea, C. (2022). Blame It on the AI? On the Moral Responsibility of Artificial Moral Advisors. *Philosophy & Technology*, *35*(2), 35. https://doi.org/10.1007/s13347-022-00529-z

De Canio, F. (2023). Consumer willingness to pay more for pro-environmental packages: The moderating role of familiarity. *Journal of Environmental Management,339*,117828. https://doi.org/10.1016/j.jenvman.2023.117828

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53*(6), 1024–1037. https://doi.org/10.1037/0022-3514.53.6.1024

Dhar, P. (2020). The carbon impact of artificial intelligence. *Nature Machine Intelligence*. *2*, 423–425.

Djeffal, C., Siewert, M. B., & Wurster, S. (2022). Role of the state and responsibility in governing artificial intelligence: a comparative analysis of AI strategies. *Journal of European Public Policy*, *29*(11), 1799–1821. https://doi.org/10.1080/13501763.2022.2094987

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data—Evolution, challenges and research agenda. *International Journal of Information Management, 48,* 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Emmet Jones, R. (2000). Measuring endorsement of the new ecological paradigm: A revised NEP scale. *Journal of Social Issues, 56*(3), 425–442. https://doi.org/10.1111/0022-4537.00176

Farrow, K., Grolleau, G., & Ibanez, L. (2017). Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics*, *140*, 1–13. https://doi.org/10.1016/j.ecolecon.2017.04.017

Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Scienc*e, *24*(1), 56–62. https://doi.org/10.1177/0956797612449177

Floridi, L., & Sanders, J.W. (2004). On the morality of artificial agents. *Minds and Machines, 14(3)*, 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Formosa, P., & Ryan, M. (2021). Making moral machines: why we need artificial moral agents. *AI and Society*, *36*(3), 839–851. https://doi.org/10.1007/s00146-020-01089-6

Gansser, O.A., & Reich, C.S. (2021). A new acceptance model for artificial intelligence with extensions to UTAUT2: An empirical study in three segments of application. *Technology in Society*, *65*, 101535. https://doi.org/10.1016/j.techsoc.2021.101535

Gifford, R., & Sussmn, R. (2012). Environmental attitudes. In S. D. Clayton (Ed.), *The Oxford handbook of environmental and conservation psychology* (pp. 65–80). Oxford University Press.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Grazzini, L., Acuti, D., & Aiello, G. (2021). Solving the puzzle of sustainable fashion consumption: The role of consumers' implicit attitudes and perceived warmth. *Journal of Cleaner Production,*287,125579. https://doi.org/10.1016/j.jclepro.2020.125579

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. The MIT Press.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, *119*,106730. https://doi.org/10.1016/j.chb.2021.106730

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814. https://doi.org/10.1037/0033-295x.108.4.814

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66. https://doi.org/10.1162/0011526042365555

Haikonen, P. O. (2007). *Robot brains: Circuits and systems for conscious machines*. John Wiley & Sons.

Hernandez, J.M., Wright, S.A., & Ferminiano Rodrigues, F. (2015). Attributes versus benefits: the role of construal levels and appeal type on the persuasiveness of marketing messages. *Journal of Advertising, 44*, 243–253. https://doi.org/10.1080/00913367.2014.967425

Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technolog*y, *11*(1), 19–29. https://doi.org/10.1007/s10676-008-9167-5

Jay, C., Yu, Y., Crawford, I., James, P., Gledson, A., Shaddick, G., Haines, R., Lannelongue, L., Lines, E., Hosking, S., & Topping, D. (2024). Prioritize environmental sustainability in use of AI and data science methods. *Nature Geoscience*, *17*(2), 106–108. https://doi.org/10.1038/s41561-023-01369-y

Johnson, S. G., & Ahn, J. (2020). Principles of moral accounting: How our intuitive moral sense balances rights and wrongs. *Cognition*, *206*, 104467. https://doi.org/10.1016/j.cognition.2020.104467

Jia, F., Soucie, K., Alisat, S., Curtin, D., & Pratt, M. (2017). Are environmental issues moral issues? Moral identity in relation to protecting the natural world. *Journal of Environmental Psychology, 52,* 104–113. https://doi.org/10.1016/j.jenvp.2017.06.004

Kelly, S., Kaye, S., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*, 101925. https://doi.org/10.1016/j.tele.2022.101925

Kneer M., Stuart M. T. (2021). Playing the blame game with robots. In Bethel C., Paiva A., Broadbent E., Feil-Seifer D., Szafi D.r (Chairs), *Companion of the 2021 ACM/IEEE international conference on human-robot interaction* (pp. 407–411). Association for Computing Machinery. https://doi.org/10.1145/3434074.3447202

Krettenauer T. (2017). Pro-environmental behavior and adolescent moral development. *Journal of Research on Adolescence*, *27*(3), 581–593. https://doi.org/10.1111/jora.12300

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181. https://doi.org/10.1023/a:1026595011371

Maninger, T., & Shank, D. B. (2022). Perceptions of violations by artificial and human actors across moral foundations. *Computers in Human Behavior Reports*, *5*, 100154. https://doi.org/10.1016/j.chbr.2021.100154

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *The Sociological Quarterly*, *52*(2), 155–194. https://doi.org/10.1111/j.1533-8525.2011.01198.x

Mert, W., Suschek-Berger, J., & Tritthart, W. (2008). *Consumer acceptance of smart appliances* (*D 5.5 of WP 5 report from Smart-A project*). Graz: Inter-University Research Centre on Technology, Work and Culture.

Monroe A. E., Dillon K. D., Malle B. F. (2014). Bringing free will down to earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*, *27*, 100–108. https://doi.org/10.1016/j.concog.2014.04.011

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, *21*(4), 18–21. https://doi.org/10.1109/MIS.2006.80

Nishant, R., Kennedy, M., & Corbett, J.(2020). Artificial intelligence for sustainability: challenges, opportunities, and a research agenda. *International Journal of Information Management*, *53*, 102104. https://doi.org/10.1016/j.ijinfomgt.2020.102104

Pal., N. R. (2020). In search of trustworthy and transparent intelligent systems with human-like cognitive and reasoning capabilities. *Frontiers in Robotics and AI*, *7*, 76. https://doi.org/10.3389/frobt.2020.00076

Park, E., Hwang, B., Ko, K., & Kim, D. (2017). Consumer acceptance analysis of the home energy management system. *Sustainability*, *9*(12), 2351. https://doi.org/10.3390/su9122351

Ray, J. L., Mende-Siedlecki, P., Gantman, A., & Van Bavel, J. J. (2021). The role of morality in social cognition. In *The Neural Basis of Mentalizing* (pp. 555-566). Springer. https://doi.org/10.1007/978-3-030-51890-5.ch28

Schwartz, D., & Loewenstein, G. (2020). Encouraging pro-environmental behaviour through green identity labelling. *Nature Sustainability*, *3*(9), 746-752. https://doi.org/10.1038/s41893-020-0543-4

Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54–63. https://doi.org/10.1145/3381831

Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine*, *296*, 114782. https://doi.org/10.1016/j.socscimed.2022.114782

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and Policy Considerations for Modern Deep Learning Research. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(09), 13693–13696. https://doi.org/10.1609/aaai.v34i09.7123

Sullins, J.P. (2006). When is a robot a moral agent? *International Review of Information Ethics*, *6*, 23–30.

Swanepoel, D. (2021). Does Artificial Intelligence Have Agency? In: Clowes, R.W., Gärtner, K., Hipólito, I. (eds) *The Mind-Technology Problem. Studies in Brain and Mind* (vol. 18, pp.88–104). Springer, Cham.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*(3), 451–471. https://doi.org/10.1037/0033-295X.109.3.451

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320-324. https://doi.org/10.1016/S1364-6613(03)00135-9

Urban, J., Bahník, Š., & Kohlová, M. B. (2023). Pro-environmental behavior triggers moral inference, not licensing by observers. *Environment and Behavior*, *55*(1-2), 74–98. https://doi.org/10.1177/00139165231163547

van Wynsberghe, A.(2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics*, *1*, 213–218. https://doi.org/10.1007/s43681-021-00043-6

van Wynsberghe, A., & Robbins, S. (2019). Critiquing the Reasons for Making Artificial Moral Agents. *Science and Engineering Ethics, 25*(3), 719–735. https://doi.org/10.1007/s11948-018-0030-8

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(4), e1507. https://doi.org/10.1002/widm.1507

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-019-14108-y

Wyss, A. M., Knoch, D., & Berger, S. (2022). When and how pro-environmental attitudes turn into behavior: The role of costs, benefits, and self-control. *Journal of Environmental Psychology*, *79*, 101748. https://doi.org/10.1016/j.jenvp.2021.101748

Yu, F. (2020). On AI and Human Beings. *Frontiers* (01), 30–36.

[喻丰.(2020).论人工智能与人之为人. 人民论坛·*学术前沿*(01), 30–36.]

Yu, F., & Xu, L.(2018). How to make an ethical intelligence? Answer from a psychological perspective. *Global Journal of Media Studies*, *5*(04), 24–42.

[喻丰,许丽颖.(2018).如何做出道德的人工智能体?——心理学的视角. *全球传媒学刊*(04), 24–42.]

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008

Unsustainability Decreases Acceptance of Environmental Artificial Intelligence

WEI Xinni[1], YU Feng[2], PENG Kaiping[1]

[1]Department of Cognitive and Psychological Science, Tsinghua University

[2] Department of Psychology, Wuhan University

Abstract Climate change and environmental issues significantly impact human health and well-being, posing substantial challenges to global sustainable development. Addressing these challenges necessitates both immediate and long-term solutions. While climate change itself does not inherently elicit a moral response, potentially hindering public engagement in climate action, artificial intelligence (AI)—encompassing robots, algorithms, and models—emerges as a promising ally. AI's ability to learn from experience, adapt to new inputs, and perform human-like tasks positions it as a critical tool in addressing environmental challenges.

However, AI presents a dual-edged impact on the environment. While it can support ecological governance and promote sustainability, its energy-intensive nature and associated carbon emissions may undermine environmental health. This study investigates the environmental implications of AI, focusing on how perceptions of AI's sustainability influence public acceptance and the underlying psychological mechanisms.

Drawing on prior research, we hypothesized that individuals tend to avoid using unsustainable AI due to perceptions of diminished morality and heightened dependency (i.e., lower perceived agency). To test this, we conducted five studies employing diverse methodologies and measures. A preliminary study used 14 words generated by ChatGPT to gauge public attitudes toward AI in environmental decision-making, revealing a generally favorable view of AI in environmental management. Studies 1a and 1b experimentally manipulated perceptions of AI's unsustainability, demonstrating that awareness of its high energy consumption and carbon emissions significantly reduces acceptance. Study 2 replicated these findings and identified morality—rather than dependency—as the mediating factor. Study 3 further revealed that pro-environmental attitudes significantly moderate the relationship between AI sustainability and its acceptance in environmental contexts.

In summary, this research highlights that while individuals are willing to collaborate with AI to address environmental challenges, their acceptance diminishes when AI is perceived as environmentally harmful. These findings underscore the critical importance of AI's sustainability in achieving global sustainable development goals.

Keywords sustainability, artificial intelligence, morality, acceptance